UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/716,202	11/19/2003	Mark D. Wasson	P68795US0	4117

136 7590 11/26/2008

JACOBSON HOLMAN PLLC
400 SEVENTH STREET N.W.
SUITE 600
WASHINGTON, DC 20004

EXAMINER	
SHAH, PARAS D	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
11/26/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/716,202	WASSON ET AL.	
	Examiner	Art Unit	
	PARAS SHAH	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *09 January 2008*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *4,5,7,9,11,12,15,16,19-21,25,41,43,45,48,49,51,52 and 56-62* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *4, 5, 7, 9, 11, 12, 15, 16, 19-21, 25, 41, 43, 45, 48, 49, 51, 52, 56-62* _____ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This communication is in response to the Arguments and Amendments filed on 08/26/2008. Claims 4, 5, 7, 9, 11, 12, 15, 16, 19-21, 25, 41, 43, 45, 48, 49, 51, 52, 56-62 are pending and have been examined, with claims 1-3, 6, 9, 10, 13, 14, 17, 18, 22-24, 26-40, 42, 44, 46, 47, 50, and 53-55 being cancelled. The Applicants' amendment and remarks have been carefully considered, but they do not place the claims in condition for allowance.

2. All previous objections and rejections directed to the Applicant's disclosure and claims not discussed in this Office Action have been withdrawn by the Examiner.

Continued Examination Under 37 CFR 1.114

3. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 08/26/2008 has been entered.

Response to Arguments

4. Applicant's arguments (pages 17-25) filed on 08/26/2008 with regard to the rejections have been fully considered but are moot in view of new grounds for rejection.

Response to Amendment

5. Applicants' amendments filed on 08/26/2008 have been fully considered. The newly amended limitations necessitate new grounds of rejection.

Specification

6. The amendment filed 08/26/2008 is objected to under 35 U.S.C. 132(a) because it introduces new matter into the disclosure. 35 U.S.C. 132(a) states that no amendment shall introduce new matter into the disclosure of the invention. The added material which is not supported by the original disclosure is as follows: In the amended Specification, paragraphs [00099] and [000115], where "FEX annotations are captured in a single view of the document expressed as inline XML" and in paragraphs [000175] and [000186], where "in a single view of the annotated document" are considered to be new matter..

Applicant is required to cancel the new matter in the reply to this Office Action.

Claim Rejections - 35 USC § 112

7. The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

8. Claims 57-59 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to

one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. Specifically, the limitation of "annotating the text represents the annotated text as a single view of the document expressed as inline XML" has been newly added subject matter, which was not defined in the Specification as originally filed.

Claim Rejections - 35 USC § 101

9. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 7, 12, 20 are rejected under 35 U.S.C. 101 because the claims appear to be directed to a software embodiment and not to a hardware embodiment, where a machine claim is directed towards a system, apparatus, or arrangement. The claim appears to be directed towards an application as stated in the published application, paragraph [0516], where the FEX tool set exists part of a larger application. There is reference that the tool set is not a free-standing application" but works as of a larger application. Thus, the claim is directed to software and is non-statutory.

10. Claims 4, 5, 8, 11, 15, 21, 25, 56-61 are rejected as being dependent upon a rejected base claim.

11. Claims 16 and 19 are rejected under 35 U.S.C. 101 because the claims do not appear to be tied to any statutory category, i.e. methods, machine, composition of matter, or article of manufacture. Instead the claims seem to be directed towards a rule-based language. Hence, the claims are directed towards non-statutory subject matter.

12. Claims 41, 43, 45, 48, 49, 51, 52, 59, and 62 are drawn to non-statutory subject matter. The stated claim falls within the statutory category of process. However, in order for the process to be statutory the process must be 1) tied to another statutory class or 2) transform underlying subject matter to a different state or thing. Neither of these requirements is met by the claim and thus the method is not a patent eligible process under 101. the claims are neither tied to another statutory class as there is no positively recited structural component to tie it to another statutory category and do not result in a transformation into a different state.

Claim Rejections - 35 USC § 103

13. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

14. Claims 4, 7, 8, 12, 15, 41, 43, 48, 49, 52, and 56-62 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov et al. ("Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov in view of Krauthammer et al. ("Representing semantic information in a linear string of text using XML", 2002), hereinafter, Krauthammer.

As to claims 7 and 41, Cunningham *et al.* teaches a fact extraction tool set for extracting information from a document, comprising:

means for breaking the text tokens (see sect 6.1, page 94, 1st paragraph, tokenizer)

means for annotating text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph) with token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph).

means for extracting facts from the annotated text (see page 104, 6.8) pattern recognition rules, (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used),wherein the pattern recognition rule comprises a pattern that describes the text of interest (see page

82, 3rd paragraph, and rule below) (e.g. From the cited portion a definition of GazLocation is given for a portion of the pattern. This is an example of a rule.), a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and an action that indicates what should be done in response to a successful matching of the pattern (see page 142, numeral2, subnumeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based functionality (see page 7, sect. 1.3.3., last two lines), and user defined matching functions.

However, Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions and tree-based functionality.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and a tree based functionality (see page 5, sect. 4.1, lines 6-8, XPATH and see page 6, left column, 1st full paragraph, last 4 lines) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by

Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8).

However, Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

As to claim 4, Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 3, above.

Furthermore, Cunningham *et al.* teaches wherein, the attributes include tokenization (see sect. 6.1), text normalization (see , part of speech tags (see sect. 6.4.), sentence boundaries (see sect. 6.3), parse trees (see page 62, sect. 4.42, last paragraph-page 63, first three lines) (e.g. It is seen that annotations

can be represented in hierarchical representation of a parse tree), semantic attribute tagging (see. sect. 6.5) and other interesting attributes of the text (see sect. 6.6).

As to claims 8 and 48, Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 3, above.

Furthermore, Cunningham *et al.* teaches wherein,

the token attributes have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute to each base token (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type).;

the constituent attributes are assigned yes-no values, where the entire pattern of each base token is considered to be a single constituent with respect to some annotation value (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text. Further, it is seen that the all of the tokens represent a pattern associated with the sentence.);

the links assign common identifiers to coreferring and other related patterns of base tokens (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal

coreference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")).

As to claim 12, Cunningham *et al.* teaches a fact extraction tool set for extracting information from a document, wherein the document includes text, comprising:

means for breaking the text tokens (see sect 6.1, page 94, 1st paragraph, tokenizer)

means for annotating text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph) with token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph)

means for associating all annotations assigned to a particular piece of text (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then

specifying an annotation based upon previous assignment), with the base tokens for that text to generate aligned annotations (e.g. This occurs when matching patterns.)

means for identifying and extracting potentially interesting pieces of information (see page 104, 6.8) in the aligned annotations by finding patterns in the attributes of the annotated text using text pattern recognition rules written in a rule based information extraction language, (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used, where the language is XML (see page 60, sect. 4.4.1, 1st paragraph), wherein the pattern recognition rule comprises a pattern that describes the text of interest (see page 82, 3rd paragraph, and rule below) (e.g. From the cited portion a definition of GazLocation is given for a portion of the pattern. This is an example of a rule.), a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and an action that indicates what should be done in response to a successful matching of the pattern (see page 142, numeral2, subnumeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based functionality (see page 7, sect. 1.3.3., last two lines), and each text pattern recognition rule queries for at least one of literal

text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)).

However, Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions and tree-based functionality.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and a tree based functionality (see page 5, sect. 4.1, lines 6-8, XPATH and see page 6, left column, 1st full paragraph, last 4 lines) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8)

However, Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

As to claims 15, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 12, above.

Furthermore, Simov does teach user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

Furthermore, Cunningham *et al.* teaches editing function to name (see page 37, sect. 2.14.2, allows name to be edited of annotations, which define the pattern to detect. The annotations are used to detect various structures in the document for extraction) and define a fragment of a pattern (see page 84, 1st and

2nd paragraph) (e.g. The label is assigned to the year based on the pattern of word in or by found in the text).

As to claim 43, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein in the annotating step the attributes include, orthographic (see sect 6.1, page 94, 1st paragraph), syntactic (see sect. 4.4.2, last paragraph), semantic (see sect. 6.5), pragmatic (see sect. 6.7.1, 1st paragraph) (e.g. The applicant refers to pragmatic as being identifying quotations, see Applicants specification, page 23, line 4) and dictionary-based attributes (see sect. 6.6.2 and see 6.2) (e.g. A table is used to determine id strings are of the same entity and the latter citation refers to names and cities).

As to claim 49, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein the means for annotating a text further comprises means for associating all annotations assigned to a particular piece of text (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous

assignment), with the base tokens for that text to generate aligned annotations (e.g. This occurs when matching patterns.).

As to claim 52, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein, text pattern recognition rule (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used) queries for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)).

As to claim 56, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, above.

Furthermore, Cunningham teaches wherein the text pattern recognition rules (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are

used, where the language is XML (see page 60, sect. 4.4.1, 1st paragraph) query for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)),

As to claims 57-59, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, 12, and 41, above.

Furthermore, Cunningham et al. teaches annotation of text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph).

Furthermore, Simov teaches wherein representation of a single view of the document expressed as inline XML (see page 6, right column, code in between 3rd paragraph, where <s> and code in between<s/>, shows the single view with annotations of parts of speech in a single view).

As to claims 60-62, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, 12, and 41, above.

Furthermore, Simov teaches wherein the means for extracting uses XPath for traversing XML-based tree representation in the annotated text (see page 7, right column, last paragraph, XPath used for extracting information).

15. Claims 5, 41, 20, 21, and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov et al. ("Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov in view of Krauthammer et al. ("Representing semantic information in a linear string of text using XML", 2002), hereinafter, Krauthammer as applied to claim 7, above and further in view of Cunningham et al. ("Gate: an architecture for development of robust HLT applications", 2002), hereinafter, Cunningham (2).

As to claim 5 and 45, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, above.

Furthermore, Cunningham teaches wherein the means for annotating the text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph) comprises, a plurality of independent annotators, wherein each of the annotators has at least one specific annotation function (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph, such as POS and semantic taggers, where each function independent of each other).

However, Cunningham *et al.* in view of Simov in view of Krauthammer do not specifically teach the user-implemented means for specifying which of the annotators to use an the order of their use.

Cunningham (2) does teach user-implemented means for specifying which of the annotators to use an the order of their use (see Figure 1, right hand-pane, where checkboxes are used for selecting annotations, and page 3, left column,

1st full paragraph, where a GUI is used for user to specify order and which processing resources to use for a specific application.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov in view of Krauthammer with user implemented means as taught by Cunningham (2) as the system of Cunningham (2) is a more detailed view of the Gate system as in Cunningham hence the purpose of combining allows information extraction (see Cunningham (2), Abstract).

As to claim 20, Cunningham *et al.* teaches wherein, a text annotation tool comprising:

means for breaking the text passage into its base tokens (see sect 6.1, page 94, 1st paragraph, tokenizer);

a plurality of independent annotators for annotating text with (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph, such as POS and semantic taggers, where each function independent of each other)with token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this

cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph).

means for associating all annotations assigned to a particular piece of text with the base tokens for that text to generate aligned annotations. (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment), with the base tokens for that text to generate aligned annotations (e.g. This is implied when matching patterns.)

However, Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions and tree-based functionality.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and a tree based functionality (see page 5, sect. 4.1, lines 6-8, XPATH and see page 6, left column, 1st full paragraph, last 4 lines) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8).

However, Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

However, Cunningham *et al.* in view of Simov in view of Krauthammer do not specifically teach the user-implemented means for specifying which of the annotators to use an the order of their use.

Cunningham (2) does teach user-implemented means for specifying which of the annotators to use an the order of their use (see Figure 1, right hand-pane,

where checkboxes are used for selecting annotations, and page 3, left column, 1st full paragraph, where a GUI is used for user to specify order and which processing resources to use for a specific application.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov in view of Krauthammer with user implemented means as taught by Cunningham (2) as the system of Cunningham (2) is a more detailed view of the Gate system as in Cunningham hence the purpose of combining allows information extraction (see Cunningham (2), Abstract).

As to claim 21, Cunningham *et al.* in view of Simov teach al of the limitations as in claim 20, above.

Furthermore, Cunningham *et al.* teaches wherein, the attributes include tokenization (see sect. 6.1), text normalization (see , part of speech tags (see sect. 6.4.), sentence boundaries (see sect. 6.3), parse trees (see page 62, sect. 4.42, last paragraph-page 63, first three lines) (e.g. It is seen that annotations can be represented in hierarchical representation of a parse tree), semantic attribute tagging (see. sect. 6.5) and other interesting attributes of the text (see sect. 6.6).

As to claims 25, Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 20, above.

Furthermore, Cunningham *et al.* teaches wherein,

the token attributes have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute to each base token (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type).;

the constituent attributes are assigned yes-no values, where the entire pattern of each base token is considered to be a single constituent with respect to some annotation value (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text. Further, it is seen that the all of the tokens represent a pattern associated with the sentence.);

where the links assign common identifiers to coreferring and other related patterns of base tokens (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal coreference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")).

16. Claims 16 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cunningham *et al.* (Developing Language Processing Component with GATE (a

User Guide), 2001-2002) in view of Simov et al. ("Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov.

As to claim 16, Cunningham *et al.* teaches a rule based information extraction language for use in identifying and extracting potentially interesting pieces of information in a text (see page 104, 6.8) the language comprising a plurality text pattern recognition rules (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used, where the language is XML (see page 60, sect. 4.4.1, 1st paragraph) that queries for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)), wherein each of the text pattern recognition rule comprises:

a pattern that describes text of interest (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used),

a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and

an action that indicates what should be done in response to a matching of the pattern (see page 142, numeral2, sub numeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based functionality (see page 7, sect. 1.3.3., last two lines); and

wherein the text pattern recognition rules use regular expression –based functionality (see page 7, sect. 1.3.3., last two lines)

However, Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions and tree-based functionality.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and a tree based functionality (see page 5, sect. 4.1, lines 6-8, XPATH and see page 6, left column, 1st full paragraph, last 4 lines) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain

information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8).

As to claims 19, Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 12, above.

Furthermore, Simov does teach user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

Furthermore, Cunningham *et al.* teaches editing function to name (see page 37, sect. 2.14.2, allows name to be edited of annotations, which define the pattern to detect. The annotations are used to detect various structures in the document for extraction) and define a fragment of a pattern (see page 84, 1st and 2nd paragraph) (e.g. The label is assigned to the year based on the pattern of word in or by found in the text).F.

17. Claims 11 and 51 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cunningham *et al.* in view of Simov in view of Krauthammer as applied to claims 12 and 41, above and further in view of Marcus *et al.* ("The PENN Treebank Annotating Predicate Argument Structure", 1994).

As to claim 11 and 51, Cunnigham *et al.* in view of Simov in view of Krauthammer teach all of the limitations as in claim 12 and 41, above.

Furthermore, Cunningham *et al.* discloses wherein the means for identifying and extracting potentially interesting pieces of information performs the further function of recognizing both true left and right constituent attributes (see sect. 6.1.1 and page 81, 1st paragraph) (e.g. It is seen that a left and right attributes are recognized by the tokeniser. Further it is admitted in the Applicant's background that many pattern recognition languages have rules that process text in left to right fashion(see Applicant's Specification, page 3, lines 2-3)) and constituent attributes (see page 63, 1st paragraph).

However, Cunningham *et al.* in view of Simov in view of Krauthammer does not specifically disclose the identification of non-contiguous attributes.

Marcus *et al.* does disclose the identification of non-contiguous attributes (see page 117, sect. 6, 2nd paragraph and example at bottom of page 117 on right hand column) (e.g. An index number is added to the label of the original constituent and allows interpretation).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction taught by Cunningham *et al.* in view of Simov in view of Krauthammer with the identification of non-contiguous attributes taught by Marcus *et al.*. The motivation to have combined the references involves the ability to represent sentences where complements of verbs occur after a sentenial level verb (see Marcus *et al.*, page 117, sect. 6, 1st paragraph), which would benefit the fact extraction tool taught by

Cunningham *et al.* in view of Simov in view of Krauthammer for recognizing discontinuous constituents.

Conclusion

18. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Solmer (US 2002/0165717) is cited to disclose a information extraction for documents by first converting the document into a standard format. Kobayashi et al. (US 2003/0007397) is cited to disclose a document processing method of tagging input data and then using tree-structured data for extraction. Masuichi et al. (US 2003/0158723) is cited to disclose a syntactic information tagging of sentences.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571)272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/P. S./
Examiner, Art Unit 2626

11/17/2008

/Patrick N. Edouard/

Supervisory Patent Examiner, Art Unit 2626